

## **RAW SEQUENCE LISTING**

**The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.**

Application Serial Number: 10/575,628  
Source: IFWP  
Date Processed by STIC: 05/04/2006

***ENTERED***



IFWP

**RAW SEQUENCE LISTING**  
PATENT APPLICATION: US/10/575,628

DATE: 05/04/2006  
TIME: 14:03:01

Input Set : A:\pto.da.txt  
Output Set: N:\CRF4\05042006\J575628.raw

4 <110> APPLICANT: CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE  
 6 <120> TITLE OF INVENTION: METHOD FOR THE DISSOCIATION OF THE EXTRACELLULAR HAEMOGLOBIN  
 7 MOLECULE OF ARENICOLA MARINA, CHARACTERIZATION OF THE PROTEIN  
 8 CHAINS CONSTITUTING SAID MOLECULE AND THE NUCLEOTIDE SEQUENCES  
 9 ENCODING SAID PROTEIN CHAINS  
 11 <130> FILE REFERENCE: WOB CNR GLOB  
 C--> 13 <140> CURRENT APPLICATION NUMBER: US/10/575,628  
 C--> 13 <141> CURRENT FILING DATE: 2006-04-13  
 13 <150> PRIOR APPLICATION NUMBER: FR 03/11992  
 14 <151> PRIOR FILING DATE: 2003-10-14  
 16 <160> NUMBER OF SEQ ID NOS: 31  
 18 <170> SOFTWARE: PatentIn version 3.1  
 20 <210> SEQ ID NO: 1  
 21 <211> LENGTH: 474  
 22 <212> TYPE: DNA  
 23 <213> ORGANISM: Arenicola marina  
 25 <220> FEATURE:  
 26 <221> NAME/KEY: CDS  
 27 <222> LOCATION: (1)..(474)  
 28 <223> OTHER INFORMATION:  
 W--> 30 <400> 1  
 31 atg aag tcc ttg gtg gtt ctg ttc gcc ctg gtg gcc atg gtg gct gca 48  
 32 Met Lys Ser Leu Val Val Leu Phe Ala Leu Val Ala Met Val Ala Ala  
 33 1 5 10 15  
 35 gag tgc ggc ccc atg cag cgc ctc ctg gtc aag acc cag tgg aac aag 96  
 36 Glu Cys Gly Pro Met Gln Arg Leu Leu Val Lys Thr Gln Trp Asn Lys  
 37 20 25 30  
 39 gtg tac ggc acc agc aag gtc agg gac qag gcc qga cac qtc ctc tgg 144  
 40 Val Tyr Gly Thr Ser Lys Val Arg Asp Glu Ala Gly His Val Leu Trp  
 41 35 40 45  
 43 aag gct att ttc gcc cag gat ccc gag acc cgg gct ctc ttc aag aga 192  
 44 Lys Ala Ile Phe Ala Gln Asp Pro Glu Thr Arg Ala Leu Phe Lys Arg  
 45 50 55 60  
 47 gtc aac ggt gac gac atc tac tct ccc gag ttc atg gct cac agc gcc 240  
 48 Val Asn Gly Asp Asp Ile Tyr Ser Pro Glu Phe Met Ala His Ser Ala  
 49 65 70 75 80  
 51 cgt gtc ttg ggt ggc ctt gac att gcc atc tcc ctc gac aac cag 288  
 52 Arg Val Leu Gly Gly Leu Asp Ile Ala Ile Ser Leu Leu Asp Asn Gln  
 53 85 90 95  
 55 gct gac ctt gac gtc gcc ctg gct cac ctt cac gtg cag cac gta gaa 336  
 56 Ala Asp Leu Asp Val Ala Leu Ala His Leu His Val Gln His Val Glu  
 57 100 105 110  
 59 agg cac atc cca acc cgc tac ttc gat ctg ttc aag aac gcc ctg atg 384

(P8-6)

## **RAW SEQUENCE LISTING**

PATENT APPLICATION: US/10/575,628

DATE: 05/04/2006

TIME: 14:03:01

Input Set : A:\pto.da.txt

Output Set: N:\CRF4\05042006\J575628.raw

60 Arg His Ile Pro Thr Arg Tyr Phe Asp Leu Phe Lys Asn Ala Leu Met  
 61 115 120 125  
 63 gag tat gcc ccc agc gcc ctg gga cgc tgc ttc gat aag acc gcc tgg 432  
 64 Glu Tyr Ala Pro Ser Ala Leu Gly Arg Cys Phe Asp Lys Thr Ala Trp  
 65 130 135 140  
 67 agc tcg tgc ttt gac gtc atc gcc aac ggc atc aag gaa tag 474  
 68 Ser Ser Cys Phe Asp Val Ile Ala Asn Gly Ile Lys Glu  
 69 145 150 155  
 72 <210> SEQ ID NO: 2  
 73 <211> LENGTH: 157  
 74 <212> TYPE: PRT  
 75 <213> ORGANISM: Arenicola marina  
 77 <400> SEQUENCE: 2  
 78 Met Lys Ser Leu Val Val Leu Phe Ala Leu Val Ala Met Val Ala Ala  
 79 1 5 10 15  
 81 Glu Cys Gly Pro Met Gln Arg Leu Leu Val Lys Thr Gln Trp Asn Lys  
 82 20 25 30  
 84 Val Tyr Gly Thr Ser Lys Val Arg Asp Glu Ala Gly His Val Leu Trp  
 85 35 40 45  
 87 Lys Ala Ile Phe Ala Gln Asp Pro Glu Thr Arg Ala Leu Phe Lys Arg  
 88 50 55 60  
 90 Val Asn Gly Asp Asp Ile Tyr Ser Pro Glu Phe Met Ala His Ser Ala  
 91 65 70 75 80  
 93 Arg Val Leu Gly Gly Leu Asp Ile Ala Ile Ser Leu Leu Asp Asn Gln  
 94 85 90 95  
 96 Ala Asp Leu Asp Val Ala Leu Ala His Leu His Val Gln His Val Glu  
 97 100 105 110  
 99 Arg His Ile Pro Thr Arg Tyr Phe Asp Leu Phe Lys Asn Ala Leu Met  
 100 115 120 125  
 102 Glu Tyr Ala Pro Ser Ala Leu Gly Arg Cys Phe Asp Lys Thr Ala Trp  
 103 130 135 140  
 105 Ser Ser Cys Phe Asp Val Ile Ala Asn Gly Ile Lys Glu  
 106 145 150 155  
 109 <210> SEQ ID NO: 3  
 110 <211> LENGTH: 477  
 111 <212> TYPE: DNA  
 112 <213> ORGANISM: Arenicola marina  
 114 <220> FEATURE:  
 115 <221> NAME/KEY: CDS  
 116 <222> LOCATION: (1)..(477)  
 117 <223> OTHER INFORMATION:  
 119 <400> 3  
 120 atg aag ttc ttg gtg gtt ctg ttc gcc ctg gtg gcc atg gtc gct gct 48  
 121 Met Lys Phe Leu Val Val Leu Phe Ala Leu Val Ala Met Val Ala Ala  
 122 1 5 10 15  
 124 gat tgt ggc ccc atg cag cgc ctc ctg gtc aag gcc cag tgg aac aag 96  
 125 Asp Cys Gly Pro Met Gln Arg Leu Leu Val Lys Ala Gln Trp Asn Lys  
 126 20 25 30  
 128 qtq tac qqc acc aqc aaq gtc agg qac qac qcc qga cac gtc ctc tgg 144

**RAW SEQUENCE LISTING**  
**PATENT APPLICATION: US/10/575,628**

**DATE: 05/04/2006**  
**TIME: 14:03:01**

**Input Set : A:\pto.da.txt**  
**Output Set: N:\CRF4\05042006\J575628.raw**

129 Val Tyr Gly Thr Ser Lys Val Arg Asp Asp Ala Gly His Val Leu Trp			
130 35	40	45	
132 aag gct atc ttc aac cag gat ggt gag acc cgc gcc ctc ttc aac aga			192
133 Lys Ala Ile Phe Asn Gln Asp Gly Glu Thr Arg Ala Leu Phe Asn Arg			
134 50	55	60	
136 gtg cac ggt gac gac atc tac tct ccc gag ttc atg gct cac agc gcc			240
137 Val His Gly Asp Asp Ile Tyr Ser Pro Glu Phe Met Ala His Ser Ala			
138 65	70	75	80
140 cgt gtc ttg ggt ggc ctt gac att gcc atc tcc ctc ctc gac aac cag			288
141 Arg Val Leu Gly Gly Leu Asp Ile Ala Ile Ser Leu Leu Asp Asn Gln			
142 85	90	95	
144 gct gag ctt gac gct gtc ctg gct cac ctc aag gag cag cac att gag			336
145 Ala Glu Leu Asp Ala Val Leu Ala His Leu Lys Glu Gln His Ile Glu			
146 100	105	110	
148 agg ggg atc cca gac cgt tac ttc gac ctg ttc aag aac gcc ctg atg			384
149 Arg Gly Ile Pro Asp Arg Tyr Phe Asp Leu Phe Lys Asn Ala Leu Met			
150 115	120	125	
152 gag ttt gcc ccc agc gcc ttg gga cgc tgc ttc ata aag gac gct tgg			432
153 Glu Phe Ala Pro Ser Ala Leu Gly Arg Cys Phe Ile Lys Asp Ala Trp			
154 130	135	140	
156 agc tca tgc ttt gac gtc att gcc aac ggc atc aag gga cag taa			477
157 Ser Ser Cys Phe Asp Val Ile Ala Asn Gly Ile Lys Gly Gln			
158 145	150	155	
161 <210> SEQ ID NO: 4			
162 <211> LENGTH: 158			
163 <212> TYPE: PRT			
164 <213> ORGANISM: Arenicola marina			
166 <400> SEQUENCE: 4			
167 Met Lys Phe Leu Val Val Leu Phe Ala Leu Val Ala Met Val Ala Ala			
168 1 5	10	15	
170 Asp Cys Gly Pro Met Gln Arg Leu Leu Val Lys Ala Gln Trp Asn Lys			
171 20	25	30	
173 Val Tyr Gly Thr Ser Lys Val Arg Asp Asp Ala Gly His Val Leu Trp			
174 35	40	45	
176 Lys Ala Ile Phe Asn Gln Asp Gly Glu Thr Arg Ala Leu Phe Asn Arg			
177 50	55	60	
179 Val His Gly Asp Asp Ile Tyr Ser Pro Glu Phe Met Ala His Ser Ala			
180 65	70	75	80
182 Arg Val Leu Gly Gly Leu Asp Ile Ala Ile Ser Leu Leu Asp Asn Gln			
183 85	90	95	
185 Ala Glu Leu Asp Ala Val Leu Ala His Leu Lys Glu Gln His Ile Glu			
186 100	105	110	
188 Arg Gly Ile Pro Asp Arg Tyr Phe Asp Leu Phe Lys Asn Ala Leu Met			
189 115	120	125	
191 Glu Phe Ala Pro Ser Ala Leu Gly Arg Cys Phe Ile Lys Asp Ala Trp			
192 130	135	140	
194 Ser Ser Cys Phe Asp Val Ile Ala Asn Gly Ile Lys Gly Gln			
195 145	150	155	
198 <210> SEQ ID NO: 5			

**RAW SEQUENCE LISTING**  
**PATENT APPLICATION: US/10/575,628**

**DATE: 05/04/2006**  
**TIME: 14:03:01**

**Input Set : A:\pto.da.txt**  
**Output Set: N:\CRF4\05042006\J575628.raw**

199 <211> LENGTH: 474  
 200 <212> TYPE: DNA  
 201 <213> ORGANISM: Arenicola marina  
 203 <220> FEATURE:  
 204 <221> NAME/KEY: CDS  
 205 <222> LOCATION: (1)..(474)  
 206 <223> OTHER INFORMATION:  
 W--> 208 <400> 5  
 209 atg aag gtc ctg atc gta ctg atg gcc tgc ttg gcc tac gtc gcc gcc 48  
 210 Met Lys Val Leu Ile Val Leu Met Ala Cys Leu Ala Tyr Val Ala Ala  
 211 1 5 10 15  
 213 gac tgc gga cct ctg cag agg ctg aag gtg aag cat cag tgg gtg cag 96  
 214 Asp Cys Gly Pro Leu Gln Arg Leu Lys Val Lys His Gln Trp Val Gln  
 215 20 25 30  
 217 gtg tac agc ggc cat ggt tac gag cgt gag gcg ttc ggc aga gag gtc 144  
 218 Val Tyr Ser Gly His Gly Tyr Glu Arg Glu Ala Phe Gly Arg Glu Val  
 219 35 40 45  
 221 ttc ctc gag atg tac aac cag gca ccc aag gcc aag gac ctc ttc acc 192  
 222 Phe Leu Glu Met Tyr Asn Gln Ala Pro Lys Ala Lys Asp Leu Phe Thr  
 223 50 55 60  
 225 agg gtc agg ggc gag aac gtc ttc tcc ccc gag ttc gga gcc cac atg 240  
 226 Arg Val Arg Gly Glu Asn Val Phe Ser Pro Glu Phe Gly Ala His Met  
 227 65 70 75 80  
 229 gtc cgt gtg ctc gga gga ctc gac atg tgc atc gct ctg ctg tcc gat 288  
 230 Val Arg Val Leu Gly Gly Leu Asp Met Cys Ile Ala Leu Ser Asp  
 231 85 90 95  
 233 gac acc gtc ctc aac gcc cag ctt gct cac ctc agc acg cag cac aag 336  
 234 Asp Thr Val Leu Asn Ala Gln Leu Ala His Leu Ser Thr Gln His Lys  
 235 100 105 110  
 237 gac cgt gga atc ccc aac gag tac ttc gat gtg atg aag gtc gcc ctc 384  
 238 Asp Arg Gly Ile Pro Asn Glu Tyr Phe Asp Val Met Lys Val Ala Leu  
 239 115 120 125  
 241 atg aag gtc ccc ggc cac gtt tca cac ttc gac ttc gat gcc tgg 432  
 242 Met Lys Val Val Pro Gly His Val Ser His Phe Asp Phe Asp Ala Trp  
 243 130 135 140  
 245 tct gcc tgc tat gac gtc atc gcc aac ggc atc aag cac taa 474  
 246 Ser Ala Cys Tyr Asp Val Ile Ala Asn Gly Ile Lys His  
 247 145 150 155  
 250 <210> SEQ ID NO: 6  
 251 <211> LENGTH: 157  
 252 <212> TYPE: PRT  
 253 <213> ORGANISM: Arenicola marina  
 255 <400> SEQUENCE: 6  
 256 Met Lys Val Leu Ile Val Leu Met Ala Cys Leu Ala Tyr Val Ala Ala  
 257 1 5 10 15  
 259 Asp Cys Gly Pro Leu Gln Arg Leu Lys Val Lys His Gln Trp Val Gln  
 260 20 25 30  
 262 Val Tyr Ser Gly His Gly Tyr Glu Arg Glu Ala Phe Gly Arg Glu Val  
 263 35 40 45

**RAW SEQUENCE LISTING**  
**PATENT APPLICATION: US/10/575,628**

**DATE: 05/04/2006**  
**TIME: 14:03:01**

**Input Set : A:\pto.da.txt**  
**Output Set: N:\CRF4\05042006\J575628.raw**

265 Phe Leu Glu Met Tyr Asn Gln Ala Pro Lys Ala Lys Asp Leu Phe Thr  
266 50 55 60  
268 Arg Val Arg Gly Glu Asn Val Phe Ser Pro Glu Phe Gly Ala His Met  
269 65 70 75 80  
271 Val Arg Val Leu Gly Gly Leu Asp Met Cys Ile Ala Leu Leu Ser Asp  
272 85 90 95  
274 Asp Thr Val Leu Asn Ala Gln Leu Ala His Leu Ser Thr Gln His Lys  
275 100 105 110  
277 Asp Arg Gly Ile Pro Asn Glu Tyr Phe Asp Val Met Lys Val Ala Leu  
278 115 120 125  
280 Met Lys Val Val Pro Gly His Val Ser His Phe Asp Phe Asp Ala Trp  
281 130 135 140  
283 Ser Ala Cys Tyr Asp Val Ile Ala Asn Gly Ile Lys His  
284 145 150 155

287 <210> SEQ ID NO: 7

288 <211> LENGTH: 498

289 <212> TYPE: DNA

290 <213> ORGANISM: Arenicola marina

292 <220> FEATURE:

293 <221> NAME/KEY: CDS

294 <222> LOCATION: (1)..(498)

295 <223> OTHER INFORMATION:

W--> 297 <400> 7

298 atg ctt cgt ttc gta gca ctc ttg gct ctg gtc ggc ctg gcc gtc tgt	48
299 Met Leu Arg Phe Val Ala Leu Leu Ala Leu Val Gly Leu Ala Val Cys	
300 1 5 10 15	
302 gac gac tgt tgt acc acc gag gac cgc aag gag gtc cag acg ctg tgg	96
303 Asp Asp Cys Cys Thr Thr Glu Asp Arg Lys Glu Val Gln Thr Leu Trp	
304 20 25 30	
306 agt gag atc tgg agt gcc cag ttc act ggt cgc cgt gtc cag gtt gcc	144
307 Ser Glu Ile Trp Ser Ala Gln Phe Thr Gly Arg Arg Val Gln Val Ala	
308 35 40 45	
310 cag gct gtg ttc gag gac ctc ttc cgc cgc gac ccc gag tcc aag aac	192
311 Gln Ala Val Phe Glu Asp Leu Phe Arg Arg Asp Pro Glu Ser Lys Asn	
312 50 55 60	
314 ctg ttc aag cgc gtc aat gtt gac gac atg aac agc ccc gaa ttc cac	240
315 Leu Phe Lys Arg Val Asn Val Asp Asp Met Asn Ser Pro Glu Phe His	
316 65 70 75 80	
318 gct cac tgc atc cgt gtt gtc aac ggt ctt gac acc gtg atc ggt ctc	288
319 Ala His Cys Ile Arg Val Val Asn Gly Leu Asp Thr Val Ile Gly Leu	
320 85 90 95	
322 ctt gac gac ccc gac acc ctg aag tcc cag ctc gag cac ttg gcc cag	336
323 Leu Asp Asp Pro Asp Thr Leu Lys Ser Gln Leu Glu His Leu Ala Gln	
324 100 105 110	
326 cag cac aag gag cgt gat ggc atc cac aag acc cac ttc gac gag atg	384
327 Gln His Lys Glu Arg Asp Gly Ile His Lys Thr His Phe Asp Glu Met	
328 115 120 125	
330 tcc cac gcc ttc ggc gcc gtc atg ccc cag gtc agc agc tgc ttc aac	432
331 Ser His Ala Phe Gly Ala Val Met Pro Gln Val Ser Ser Cys Phe Asn	

RAW SEQUENCE LISTING ERROR SUMMARY  
PATENT APPLICATION: US/10/575,628

DATE: 05/04/2006  
TIME: 14:03:02

Input Set : A:\pto.da.txt  
Output Set: N:\CRF4\05042006\J575628.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:21; N Pos. 9,12  
Seq#:23; N Pos. 6,12  
Seq#:24; N Pos. 6,12  
Seq#:26; N Pos. 6,15  
Seq#:27; N Pos. 15,18  
Seq#:28; N Pos. 4,7  
Seq#:29; N Pos. 2,8,11,14  
Seq#:30; N Pos. 4,7  
Seq#:31; N Pos. 3,6

**VERIFICATION SUMMARY**  
PATENT APPLICATION: US/10/575,628

DATE: 05/04/2006  
TIME: 14:03:02

Input Set : A:\pto.da.txt  
Output Set: N:\CRF4\05042006\J575628.raw

L:13 M:270 C: Current Application Number differs, Replaced Current Application No  
L:13 M:271 C: Current Filing Date differs, Replaced Current Filing Date  
L:30 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:1,Line#:28  
L:119 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:3,Line#:117  
L:208 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:5,Line#:206  
L:297 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:7,Line#:295  
L:393 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:9,Line#:391  
L:491 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:11,Line#:489  
L:625 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:13,Line#:623  
L:700 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:15,Line#:698  
L:762 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:17,Line#:760  
L:839 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:19,Line#:837  
L:932 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:21 after pos.:0  
L:966 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:23 after pos.:0  
L:988 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:24 after pos.:0  
L:1022 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:26 after pos.:0  
L:1044 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:27 after pos.:0  
L:1066 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:28 after pos.:0  
L:1099 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:29 after pos.:0  
L:1121 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:30 after pos.:0  
L:1143 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:31 after pos.:0